Active Sentiment Domain Adaptation

Fangzhao Wu[†], **Yongfeng Huang**[†,*], **and Jun Yan**[‡]
[†]Department of Electronic Engineering, Tsinghua University
[‡]Microsoft Research Asia, Beijing, China
wufangzhao@gmail.com, yfhuang@tsinghua.edu.cn, junyan@microsoft.com

Abstract

Domain adaptation is an important technology to handle domain dependence problem in sentiment analysis field. Existing methods usually rely on sentiment classifiers trained in source domains. However, their performance may heavily decline if the distributions of sentiment features in source and target domains have significant difference. In this paper, we propose an active sentiment domain adaptation approach to handle this problem. Instead of the source domain sentiment classifiers, our approach adapts the general-purpose sentiment lexicons to target domain with the help of a small number of labeled samples which are selected and annotated in an active learning mode, as well as the domain-specific sentiment similarities among words mined from unlabeled samples of target domain. A unified model is proposed to fuse different types of sentiment information and train sentiment classifier for target domain. Extensive experiments on benchmark datasets show that our approach can train accurate sentiment classifier with less labeled samples.

1 Introduction

Sentiment classification is widely known as a domain-dependent problem (Liu, 2012; Pang and Lee, 2008; Blitzer et al., 2007; Pan et al., 2010). This is because different domains usually have many different sentiment expressions. For example, "lengthy" and "boring" are popularly used in *Book* domain to express negative sentiment. However, they are rare in *Kitchen appliance* domain. Moreover, the same word or phrase may convey different sentiments in different domains. For instance, "unpredictable" is frequently used to express positive sentiment in *Movie* domain (e.g., "The plot of this movie is fun and unpredictable"). However, it tends to be used as a negative word in *Kitchen appliance* domain (e.g., "Even holding heat is unpredictable. It is just terrible!"). Thus, every domain has many domain-specific sentiment expressions, which cannot be captured by other domains. The performance of directly applying a general sentiment classifier or a sentiment classifier trained in other domains to target domain is usually suboptimal.

Since there are a large number of domains in user-generated content, it is impractical to manually annotate enough samples for each domain to train an accurate domain-specific sentiment classifier. Thus, sentiment domain adaptation, which transfers the sentiment classifier trained in a source domain with sufficient labeled data to a target domain with no or scarce labeled data, has been widely studied (Blitzer et al., 2007; Pan et al., 2010; He et al., 2011; Glorot et al., 2011). Existing sentiment domain adaptation methods are mainly based on transfer learning techniques. Many of them try to learn a new feature representation to augment or replace the original feature space in order to reduce the gap of sentiment feature distributions between source and target domains (Pan et al., 2010; Glorot et al., 2011). For example, Blitzer et al. (2007) proposed to learn a latent representation for domain-specific words from both source and target domains by using pivot features as bridge. The advantage of these methods is that no labeled data in target domain is needed. However, when the distributions of sentiment features in source and target domains have significant difference, the performance of domain adaptation will heavily decline (Li et al., 2013). In some cases, the performance of adaptation is even lower than

*Corresponding author.

Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, pages 1701–1711
Vancouver, Canada, July 30 - August 4, 2017.
https://doi.org/10.18653/v1/P17-1156

that without adaptation, which is usually known as *negative transfer* (Pan and Yang, 2010).

In this paper, we propose an active sentiment domain adaptation approach to handle this problem by incorporating both general sentiment information and a small number of actively selected labeled samples from target domain. More specifically, in our approach the general sentiment information extracted from sentiment lexicons is adapted to target domain using domain-specific sentiment similarities among words. The general sentiment information is regarded as a "background" domain to transfer. The word similarities are extracted from unlabeled samples of target domain using both syntactic rules and co-occurrence patterns. Then we actively select and annotate a small number of informative samples from target domain in an active learning manner. These labeled samples are incorporated into our approach to improve the performance of sentiment domain adaptation. A unified model is proposed to incorporate different types of sentiment information to train sentiment classifier for target domain. Extensive experiments were conducted on benchmark datasets. The experimental results show that our approach can train accurate sentiment classifiers and reduce the manual annotation effort.

2 Related Work

2.1 Sentiment Domain Adaptation

Sentiment classification is well known as a highly domain-dependent task, and domain adaptation is widely studied in sentiment analysis field to handle this problem (Blitzer et al., 2007; Pan et al., 2010; He et al., 2011; Glorot et al., 2011). Existing sentiment domain adaptation methods are mainly based on transfer learning technique (Pan and Yang, 2010), where sentiment classifiers are trained in one or multiple source domains with sufficient labeled samples, and then applied to target domain where there is no or only scarce labeled samples. In order to reduce the gap of sentiment feature distributions between source and target domains, many sentiment domain adaptation methods try to learn a new feature representation to augment or replace the original feature space. For example, Pan et al. (2010) proposed a sentiment domain adaptation method based on spectral feature alignment (*SFA*) algorithm. They first manually selected several domain-independent features and computed the associations between domain-specific features and domain-independent features. After that they built a bipartite graph where domain-independent and domain-specific features were regarded as two types of nodes. Then domain-specific features were grouped into several clusters using spectral clustering algorithm. These clusters were used to augment the original feature representations. Glorot et al. (2011) proposed a sentiment domain adaptation method based on a deep learning technique, i.e., Stacked Denoising Autoencoders. They learned the parameters of neural networks using unlabeled samples from both source and target domains, and used the hidden nodes of the neural networks as the latent feature representations of both domains. Then they trained sentiment classifiers using source domain labeled data in this new feature space and applied it to target domain. The advantage of these sentiment domain adaptation methods is that they do not rely on the labeled data in target domain. However, they have a common shortcoming, i.e., when the distributions of sentiment features in source and target domains have significant difference, the performance of domain adaptation will heavily decline (Li et al., 2013). In some cases, *negative transfer* may happen (Blitzer et al., 2007; Li et al., 2013), which means the performance of adaptation is worse than that without adaptation (Pan and Yang, 2010). Different from many existing sentiment domain adaptation methods, in our approach we adapt the general sentiment information in sentiment lexicons to target domain with the help of a small number of labeled samples which are selected and annotated in an active learning mode. Since the sentiment words in general-purpose sentiment lexicons usually convey consistent sentiment polarities in different domains, and the actively selected labeled samples contain rich domain-specific sentiment information of target domain, our approach can effectively reduce the risk of negative transfer.

The usefulness of labeled samples from target domain in sentiment domain adaptation has been observed by previous research works (Choi and Cardie, 2009; Chen et al., 2011; Li et al., 2013; Wu et al., 2016). For example, Choi and Cardie (2009) proposed to adapt a sentiment lexicon to a specific domain by exploiting both the relations among words which co-occur in the same sentiment expressions and the relations between words and labeled sentiment expressions. However, the

labeled samples used in these methods are randomly selected, while in our approach we actively select informative samples from target domain to annotate. Thus, our approach has the potential to reduce the manual annotation effort.

2.2 Active Learning

Active learning is a useful technique in scenarios where unlabeled data is abundant but their labels are difficult or expensive to obtain (Tong and Koller, 2002; Settles, 2010). By actively selecting informative samples to label, active learning can effectively reduce the annotation effort, and improve the classification performance with limited budget (Li et al., 2012). An important problem in active learning is how to evaluate the informativeness of unlabeled samples (Fu et al., 2013). Different methods have been applied to select informative samples, such as uncertainty sampling (Zhu et al., 2010; Yang et al., 2015), query-by-committee (Freund et al., 1997; Li et al., 2013) and so on. In our approach, uncertainty combined with density is used to measure the informativeness of samples. A major difference between our approach and existing active learning methods is that in existing methods the parameters of the initial classifier are either initialized as zero (Cesa-Bianchi et al., 2006) or learned from a set of randomly selected samples (Settles, 2010). In contrast, the initial sentiment classifier in our approach is constructed by adapting the general sentiment information to target domain via the domain-specific sentiment similarities among words.

There are a few works that apply active learning methods to sentiment domain adaptation task (Rai et al., 2010; Li et al., 2013). For example, Rai et al. (2010) proposed an online active learning algorithm for sentiment domain adaptation. They started with a sentiment classifier trained on the labeled samples of a source domain. Then they sequentially selected informative samples in target domain to annotate with a probability positively related to classification uncertainty. The newly annotated samples were used to update the sentiment classifier in an online learning manner. Li et al. (2013) proposed another active learning method for cross-domain sentiment classification. In their method they trained two sentiment classifiers, one on the labeled samples of source domain, and the other one on the labeled samples of target domain. Then query-by-committee strategy was used to select the informative instances from target domain. Different from these methods, our approach does not rely on the labeled data of source domains. Instead, in our approach the general sentiment information in sentiment lexicons is actively adapted to target domain, which usually has better generalization ability in various domains than the sentiment classifier trained in a source domain. In addition, our approach can incorporate the domain-specific sentiment similarities among words mined from unlabeled samples of target domain, which are not considered in these methods.

3 Active Sentiment Domain Adaptation

3.1 Notations

First we introduce several notations that will be used in remaining part of this paper. Denote the general sentiment information extracted from a general-purpose sentiment lexicon as $\mathbf{p} \in \mathbb{R}^{D\times 1}$, where D is the vocabulary size. If the i_{th} word is labeled as positive (or negative) in the sentiment lexicon, then $p_i = +1$ (or $p_i = -1$). Otherwise, $p_i = 0$. Following many previous works in sentiment classification field (Blitzer et al., 2007; Pan et al., 2010), here we select linear classifier as sentiment classifier, and denote the linear classification model as $\mathbf{w} \in \mathbb{R}^{D\times 1}$. We use $f(\mathbf{x}_i, y_i, \mathbf{w})$ to represent the loss of classifying the i_{th} labeled sample in target domain under the classification model $\mathbf{w}$, where f is the classification loss function, $\mathbf{x}_i \in \mathbb{R}^{D\times 1}$ is the feature vector of this sample and y_i is its sentiment label. In this paper we focus on binary sentiment classification and $y_i \in \{+1, -1\}$. In addition, we select log loss for f. Thus, $f(\mathbf{x}_i, y_i, \mathbf{w}) = \log(1 + \exp(-y_i \mathbf{w}^T \mathbf{x}_i))$. Besides, we use $\mathbf{S} \in \mathbb{R}^{D\times D}$ to represent the sentiment similarities among words extracted from unlabeled samples of target domain.

3.2 Domain-Specific Sentiment Similarities

Next we introduce the extraction of domain-specific sentiment similarities among words from unlabeled samples of target domain. Two types of similarities are extracted in this paper. The first one is based on syntactic rules, which is inspired by (Hatzivassiloglou and McKeown, 1997; Huang et al., 2014; Wu and Huang, 2016). If two words have the same POS-tag such as adjective, verb, and adverb, and they are connected by coordinating conjunction "and" in the same sentence, then we regard they convey the same sentiment polarity. In

addition, if two words are connected by adversative conjunction "but" and have the same POS-tag, then they are assumed to have opposite sentiment polarities. Denote $\mathbf{S}^r \in \mathbb{R}^{D\times D}$ as the sentiment similarities extracted from unlabeled samples according to syntactic rules, and the similarity score between words i and j is defined as:

$$S^r_{i,j} = \frac{N^s_{i,j} - N^o_{i,j}}{N^s_{i,j} + N^o_{i,j} + \alpha_1}, \tag{1}$$

where $N^s_{i,j}$ and $N^o_{i,j}$ are the frequencies of words i and j having the same or opposite sentiments respectively according to the syntactic rules, and α_1 is a positive smoothing factor. If two words have much higher frequency of sharing the same sentiment than opposite sentiments, then they will have a larger positive sentiment similarity score. Note that $S^r_{i,j}$ can be negative according to Eq. (1). Here we focus on sentiment similarity rather than dissimilarity, and set all the negative values in $\mathbf{S}^r$ to zero. The range of $S^r_{i,j}$ is $[0, 1]$.

The second type of sentiment similarities are extracted according to the co-occurrence patterns among words. It is inspired by the observation that words frequently co-occurring with each other not only have a high probability to have similar semantics, but also tend to share similar sentiments (Turney, 2002; Velikovich et al., 2010; Yogatama and Smith, 2014; Tang et al., 2015; Hamilton et al., 2016). In this paper, we compute the co-occurrence between words in the context of document. Denote $\mathcal{D}$ as the set of all documents, and N^i_d as the frequency of word i appearing in document d. Then, the sentiment similarity score between words i and j based on their co-occurrence patterns is defined as:

$$S^c_{i,j} = \frac{\sum_{d\in\mathcal{D}} \min\{N^i_d, N^j_d\}}{\sum_{d\in\mathcal{D}} \max\{N^i_d, N^j_d\} + \alpha_2}, \tag{2}$$

where α_2 is a positive smoothing parameter. If two words frequently co-occur with each other in many documents, then they will have a high sentiment similarity score according to Eq. (2). The range of $S^c_{i,j}$ is also $[0, 1]$. Denote $\mathbf{S}^c \in \mathbb{R}^{D\times D}$ as the set of all sentiment similarities extracted according to co-occurrence patterns.

The sentiment similarities extracted according to syntactic rules are usually of high accuracy. However, their coverage is limited, because the word pairs detected by these syntactic rules are sparse. In contrast, the coverage of sentiment similarities extracted from co-occurrence patterns is quite wide because document is a long context, while their accuracies are not as high as the similarities based on syntactic rules. Thus, we propose to combine these two types of sentiment similarities to obtain a balance between accuracy and coverage. Denote $\mathbf{S} \in \mathbb{R}^{D\times D}$ as the final sentiment similarities among words, and $S_{i,j} = \theta S^r_{i,j} + (1-\theta) S^c_{i,j}$, where $\theta \in [0, 1]$ is the combination coefficient. In this paper we set θ to 0.5, which means that we regard these two types of sentiment similarities as equally important.

3.3 Initial Sentiment Classifier Construction

In this section, we introduce the construction of the initial sentiment classifier to start the active learning process. Existing active learning methods usually randomly select a set of unlabeled samples to annotate and then train the initial classifier on them (Settles, 2010). However, these randomly selected samples may be redundant and not informative enough. In this paper, we propose to build the initial sentiment classifier by adapting the general sentiment information to target domain via domain-specific sentiment similarities as follows:

$$\mathbf{w}_0 = \arg\min_{\mathbf{w}} -\sum_{i=1}^{D} p_i w_i + \alpha \sum_{i=1}^{D}\sum_{j\neq i} S_{i,j}(w_i - w_j)^2, \tag{3}$$

where $\mathbf{w}_0 \in \mathbb{R}^{D\times 1}$ is the initial sentiment classifier, α is a positive regularization coefficient, p_i is the prior sentiment polarity of word i in sentiment lexicons, and $S_{i,j}$ is the sentiment similarity score between words i and j. Eq. (3) is motivated by (Bengio et al., 2006), and the quadratic cost criterion is equivalent to label propagation. In Eq. (3), $-\sum_{i=1}^{D} p_i w_i$ means that if a word i is labeled as a positive (or negative) word in a general-purpose sentiment lexicon, i.e., $p_i > 0$ (or $p_i < 0$), then we constrain that its sentiment weight in the sentiment classifier is also positive (or negative). Otherwise, a penalty will be added to the objective function. In addition, $\sum_{i=1}^{D}\sum_{j\neq i} S_{i,j}(w_i - w_j)^2$ represents that if two words share high sentiment similarity, then we constrain they have similar sentiment weights in sentiment classifier. For example, if we find that "great" and "easy" have high sentiment similarities in *Kitchen appliances* domain (e.g., "This is a great pan and easy to wash"), and "great" is a positive sentiment word in many sentiment lexicons, then we can infer that "easy" may also be a positive sentiment word in this domain by propagating the sentiment information from

"great" to "easy". In this way, the general sentiment information can be adapted to many domain-specific sentiment expressions in target domain.

3.4 Query Strategy

Active learning methods iteratively select the most informative instances to label and add them to the training set (Settles, 2010). Thus, an important issue in these methods is how to measure the informativeness of unlabeled samples. In this paper, we select classification uncertainty as the informativeness measure, which has been proven effective in many active learning methods (Zhu et al., 2010; Yang et al., 2015). Since we focus on binary sentiment classification and the classification loss function is log loss, the classification uncertainty of an unlabeled instance $\mathbf{x}$ is defined as:

$$U(\mathbf{x}) = 1 - \left| 1 - \frac{2}{1 + \exp(-\mathbf{w}^T \mathbf{x})} \right|, \tag{4}$$

where $\mathbf{w}$ is the linear sentiment classification model. The range of $U(\mathbf{x})$ is $[0, 1]$. If $|\mathbf{w}^T \mathbf{x}|$ is large, which means that current sentiment classifier is confident in classifying this instance, then the uncertainty of $\mathbf{x}$ (i.e., $U(\mathbf{x})$) will be low. If $|\mathbf{w}^T \mathbf{x}|$ is close to 0, then the sentiment classifier is very uncertain about this instance, probably because the sentiment expressions in this instance are not covered by current sentiment classifier, and the uncertainty of the instance $\mathbf{x}$ will be high. In this case, annotating this instance and adding it to the training set are beneficial, because it can provide the information of unknown sentiment expressions and has the potential to quickly improve the quality of target domain sentiment classifier.

However, many researchers have found that unlabeled instances with high uncertainties can be outliers, whose labels are useless and even misleading (Settles, 2010; Zhu et al., 2010). Thus, here we combine uncertainty with representativeness to avoid outliers. Density is proven to be an effective measure of representativeness in active learning methods (Zhu et al., 2010; Hajmohammadi et al., 2015). Here we use the k-nearest neighbour based density proposed by Zhu et al. (2010) as the representativeness measure, which is formulated as:

$$D(\mathbf{x}) = \frac{1}{k} \sum_{\mathbf{x}_i \in \mathcal{N}(\mathbf{x})} \frac{\mathbf{x}^T \mathbf{x}_i}{\|\mathbf{x}\|_2 \cdot \|\mathbf{x}_i\|_2}, \tag{5}$$

where $\mathcal{N}(\mathbf{x})$ is the set of k most similar instances of $\mathbf{x}$. The final informativeness score of an unlabeled sample is a linear combination of uncertainty and density which is formulated as follows:

$$I(\mathbf{x}) = \eta(t) U(\mathbf{x}) + (1 - \eta(t)) D(\mathbf{x}), \tag{6}$$

where $\eta(t) \in [0, 1]$ is the combination coefficient at the t_{th} iteration. In this paper, we select a monotonically increasing function for $\eta(t)$, i.e., $\eta(t) = \frac{1}{1+\exp(2-\frac{4t}{T})}$, where T is the total number of iterations. It means that at initial iterations we put more emphasis on instances with high representativeness, because the initial sentiment classifier built by adapting the general sentiment information via the domain-specific sentiment similarities is relatively weak, and we prefer to select instances with more popular sentiment expressions to annotate. As more and more labeled samples are added to the training set and the sentiment classifier becomes stronger, we gradually focus on more difficult instances, i.e., those having higher classification uncertainty scores.

3.5 Active Domain Adaptation

Based on previous discussions, in this section we introduce the complete procedure of our active sentiment domain adaptation (*ASDA*) approach. Different from existing sentiment domain adaptation methods which rely on the sentiment classifier trained in source domains to transfer, in our approach we regard the general sentiment information in sentiment lexicons as the "background" domain and adapt it to target domain with the help of a small number of labeled samples which are selected and annotated in an active learning mode. First, we build an initial sentiment classifier according to Eq. (3) by adapting the general sentiment information to target domain using the domain-specific sentiment similarities among words mined from unlabeled samples of target domain. Second, we compute the density of each unlabeled sample in $\mathcal{U}$ according to Eq. (5). Then we repeat following steps until the annotation budget has run out. First, we compute the uncertainty of each unlabeled sample in $\mathcal{U}$ according to Eq. (4), and further we compute their informativeness by combining uncertainty with density according to Eq. (6). Next, we select the unlabeled sample with the highest informativeness from $\mathcal{U}$ and manually annotate its sentiment polarity. Then we add it to the set of labeled samples $\mathcal{L}$ and remove it from $\mathcal{U}$. After that we retrain the sentiment classifier for target domain based on the general sentiment information $\mathbf{p}$, the labeled samples $\mathcal{L}$, and the domain-

specific sentiment similarities $\mathbf{S}$ as follows:

$$\arg\min_{\mathbf{w}} -\sum_{i=1}^{D} p_i w_i + \alpha \sum_{i=1}^{D} \sum_{j \neq i} S_{i,j}(w_i - w_j)^2 + \beta \sum_{\mathbf{x}_i \in \mathcal{L}} \log(1 + \exp(-y_i \mathbf{w}^T \mathbf{x}_i)) + \lambda \|\mathbf{w}\|_2^2, \quad (7)$$

where α, β, and λ are nonnegative coefficients. By the term $-\sum_{i=1}^{D} p_i w_i$ we constrain that the target domain sentiment classifier learned by our approach is consistent with the general sentiment information. Through this way, the general sentiment information extracted from sentiment lexicons can be adapted to target domain. The term $\sum_{i=1}^{D} \sum_{j \neq i} S_{i,j}(w_i - w_j)^2$ is motivated by label propagation (Bengio et al., 2006). If two words tend to have high sentiment similarity with each other according to many unlabeled samples of target domain, then we constrain that their sentiment weights in the target domain sentiment classifier are also similar. The term $\sum_{\mathbf{x}_i \in \mathcal{L}} \log(1 + \exp(-y_i \mathbf{w}^T \mathbf{x}_i))$ means that we hope to minimize the empirical classification loss on labeled samples of target domain. By this term the sentiment information in the labeled samples is incorporated into the learning of target domain sentiment classifier. The L_2-norm regularization term is introduced to control model complexity. The sentiment classifier trained in Eq. (7) is further used at the next iteration of active sentiment domain adaptation until all the budget of manual annotation has been used. Then we obtain the final sentiment classifier of target domain. The complete algorithm of our active sentiment domain adaptation (*ASDA*) approach is summarized in Algorithm 1.

Algorithm 1 Active sentiment domain adaptation.

1: **Input:** The set of unlabeled samples $\mathcal{U}$, the general sentiment information $\mathbf{p}$, the domain-specific sentiment similarities $\mathbf{S}$, and the total annotation budget N.
2: **Output:** Target domain sentiment classifier $\mathbf{w}$.
3: Train the initial sentiment classifier $\mathbf{w}_0$ (Eq. (3)).
4: Compute the density of each sample $\mathbf{x}_i$ in $\mathcal{U}$ (Eq. (5)).
5: Initialize the set of labeled samples $\mathcal{L} = \emptyset$, the iteration number $t = 0$, and the sentiment classifier $\mathbf{w} = \mathbf{w}_0$.
6: **while** $t < N$ **do**
7: $t = t + 1$.
8: Compute the uncertainty score of each sample $\mathbf{x}_i$ in $\mathcal{U}$ (Eq. (4)).
9: Compute the informativeness score of each sample $\mathbf{x}_i$ in $\mathcal{U}$ (Eq. (6)).
10: Select $\mathbf{x}^*$ from $\mathcal{U}$ which has the highest informativeness score.
11: Annotate $\mathbf{x}^*$ and obtain its sentiment label y.
12: $\mathcal{L} = \mathcal{L} + \{\mathbf{x}^*, y\}$, $\mathcal{U} = \mathcal{U} - \mathbf{x}^*$.
13: Update sentiment classifier $\mathbf{w}$ according to Eq. (7).
14: **end while**

4 Experiments

4.1 Datasets

The dataset used in our experiments is the Amazon product review dataset[1] collected by Blitzer et al. (2007), which is widely used in sentiment analysis and domain adaptation research (Pan et al., 2010; Bollegala et al., 2011). This dataset contains product reviews in four domains, i.e., *Book*, *DVD*, *Electronics*, and *Kitchen appliances*. In each domain, 1,000 positive and 1,000 negative reviews as well as a large number of unlabeled samples are included. The detailed statistics of this dataset are summarized in Table 1.

	Book	*DVD*	*Electronics*	*Kitchen*
positive	1,000	1,000	1,000	1,000
negative	1,000	1,000	1,000	1,000
unlabeled	973,194	122,438	21,009	17,856

Table 1: The statistics of the Amazon dataset.

Following many previous works (Blitzer et al., 2007; Bollegala et al., 2011), unigrams and bigrams were used to build feature vectors in our experiments. We randomly split the labeled samples in each domain into two parts with equal size. The first part was used as test data, and the second part was used as the pool of "unlabeled" samples to perform active learning. The general sentiment information was extracted from Bing Liu's sentiment lexicon[2] (Hu and Liu, 2004), which is one of the state-of-the-art general-purpose sentiment lexicons. The domain-specific sentiment similarities among words were extracted from the large-scale unlabeled samples. The total number of samples actively selected by our approach to annotate was set to 100. The values of α, β, and λ were set to 0.1, 1, and 1 respectively. We repeated each experiment 10 times independently and the average results were reported.

4.2 Algorithm Effectiveness

First we conducted several experiments to explore the effectiveness of our active sentiment domain adaptation (*ASDA*) approach. We hope to answer two questions via these experiments: 1) whether the domain-specific sentiment similarities among words mined from unlabeled samples of target

[1] https://www.cs.jhu.edu/~mdredze/datasets/sentiment/

[2] https://www.cs.uic.edu/liub/FBS/sentiment-analysis.html

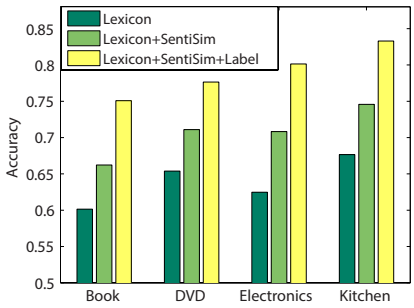


Figure 1: The performance of our approach with different combinations of sentiment information. *Lexicon*, *SentiSim*, and *Label* represent the general-purpose sentiment lexicon, the domain-specific sentiment similarities among words, and a small number of actively selected and annotated samples in target domain respectively.

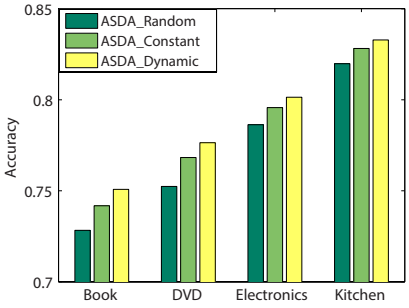


Figure 2: The performance of our approach with labeled samples selected by different strategies.

domain are useful for adapting the general sentiment information to target domain; 2) whether a small number of samples which are actively selected and annotated in target domain can help improve the domain adaptation performance. In our experiments, we implemented different versions of our *ASDA* approach using different combinations of sentiment information. The first one is *Lexicon*, which means only using the general sentiment information and no domain adaptation is conducted. It serves as a baseline. The second one is *Lexicon+SentiSim*, which means adapting general sentiment information to target domain using domain-specific sentiment similarities, but labeled samples of target domain are not incorporated. The third one is *Lexicon+SentiSim+Label*, which is the complete *ASDA* approach. The experimental results are summarized in Fig. 1.

According to Fig. 1, the performance of *Lexicon* is suboptimal. This is because the general sentiment lexicons cannot capture the domain-specific sentiment expressions in target domain (Choi and Cardie, 2009). *Lexicon+SentiSim* performs significantly better than *Lexicon*, which validates that the sentiment similarities among words extracted from unlabeled samples of target domain contain rich domain-specific sentiment information, and can help propagate the general sentiment information to many domain-specific sentiment expressions. Besides, after incorporating a small number of labeled samples which are actively selected and annotated by our approach in an active learning mode, the performance of our sentiment domain adaptation approach is significantly improved. This is because although these labeled samples are in limited size and cannot cover all the sentiment expressions in target domain, they can provide sentiment information of popular domain-specific sentiment expressions, which can be propagated to other sentiment expressions in target domain during the domain adaptation process. Thus, above experimental results validate the effectiveness of our approach.

We also conducted several experiments to verify the advantage of the actively selected samples over randomly selected samples and validate the effectiveness of our active learning algorithm. We also compared the dynamic weighting scheme for combining uncertainty and density with the constant weighting scheme. The experimental results are summarized in Fig. 2. According to Fig. 2, our approach with actively selected samples performs better than that with randomly selected samples. It indicates that these actively selected samples are more informative than randomly selected samples for sentiment domain adaptation. In addition, our approach with dynamic weighting scheme in combining uncertainty and density outperforms that with constant weighting scheme, which implies that it is beneficial to emphasize representative samples at initial iterations and gradually focus on difficult samples at later iterations. Thus, the experimental results validate the effectiveness of our active learning algorithm.

4.3 Performance Evaluation

In this section we conducted experiments to evaluate the performance of our approach by comparing it with several baseline methods. The methods to be compared include: 1) *MPQA* and *BingLiu*, using two state-of-the-art sentiment lexicons,

i.e., MPQA (Wilson et al., 2005) and Bing Liu's lexicon (Hu and Liu, 2004) for sentiment classification following the suggestions in (Hu and Liu, 2004); 2) *SVM*, *LS*, and *LR*, three popular supervised sentiment classification methods, i.e., support vector machine (Pang et al., 2002), least squares (Hu et al., 2013) and logistic regression (Wu et al., 2015); 3) *ZIAL*, the zero initialized active learning method (Cesa-Bianchi et al., 2006); 4) *LIAL*, the active learning method initialized by randomly selected labeled data (Settles, 2010); 5) *SCL* and *SFA*, two famous sentiment domain adaptation methods proposed in (Blitzer et al., 2007) and (Pan et al., 2010) respectively; 6) *ILP*, adapting sentiment lexicons to target domain via integer linear programming (Choi and Cardie, 2009); 7) *AODA*, the active online domain adaptation method (Rai et al., 2010); 8) *ALCD*, the active learning method for cross-domain sentiment classification (Li et al., 2013); 9) *ASDA*, our active sentiment domain adaptation approach. For above methods, if labeled target domain samples are needed in training, the number of labeled samples was set to 100, and if source domain labeled samples are needed in training, the number of labeled samples was set to 1,000. The parameters in baseline methods were tuned via cross-validation. The experimental results are summarized in Table 2.

	Book		*DVD*		*Electronics*		*Kitchen*	
	Acc	Fscore	Acc	Fscore	Acc	Fscore	Acc	Fscore
MPQA	0.5953	0.5673	0.6149	0.5936	0.6150	0.6070	0.6392	0.6258
BingLiu	0.6015	0.6048	0.6539	0.6604	0.6248	0.6320	0.6765	0.6930
SVM	0.6580	0.6511	0.6688	0.6652	0.7138	0.7129	0.7386	0.7412
LS	0.6543	0.6542	0.6692	0.6687	0.7194	0.7185	0.7479	0.7465
LR	0.6606	0.6582	0.6774	0.6742	0.7257	0.7226	0.7492	0.7480
RIAL	0.6693	0.6663	0.6850	0.6821	0.7310	0.7299	0.7574	0.7568
LIAL	0.6756	0.6731	0.6866	0.6838	0.7374	0.7360	0.7599	0.7595
SCL	0.7233	0.7201	0.7469	0.7438	0.7768	0.7730	0.8099	0.8095
SFA	0.7307	0.7285	0.7513	0.7485	0.7846	0.7812	0.8174	0.8153
ILP	0.6942	0.6931	0.7153	0.7124	0.7463	0.7445	0.7793	0.7768
AODA	0.6928	0.6912	0.7172	0.7165	0.7518	0.7512	0.7698	0.7690
ALCD	0.7237	0.7221	0.7369	0.7364	0.7768	0.7788	0.7979	0.7970
ASDA	0.7508	0.7501	0.7764	0.7759	0.8014	0.8011	0.8329	0.8328

Table 2: Sentiment classification performance of different methods in different domains. *Acc* and *Fscore* represent accuracy and macro-averaged Fscore respectively.

According to Table 2, the performance of directly applying sentiment lexicons to target domain is suboptimal. This is because there are many domain-specific sentiment expressions that are not covered by these general-purpose sentiment lexicons (Choi and Cardie, 2009). In addition, the performance of supervised sentiment classification methods such as *SVM*, *LS*, and *LR* is also

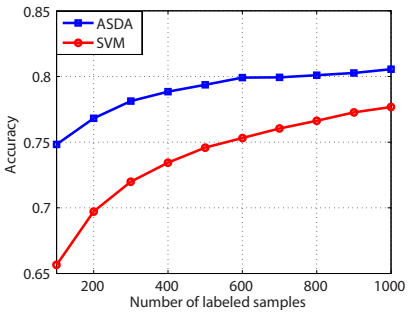
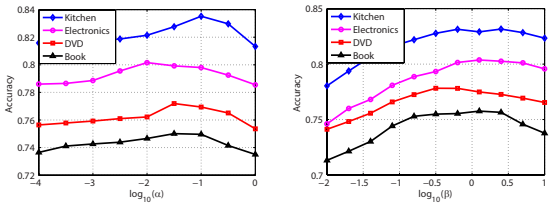


Figure 3: The performance of *ASDA* and *SVM* with different numbers of labeled samples.

limited, because the labeled samples for training are extremely scarce. The active learning methods such as *ZIAL* (Cesa-Bianchi et al., 2006) and *LIAL* (Settles, 2010) perform relatively better, because they can actively select informative samples to annotate and learn. Our approach can outperform both of them. This is because besides the labeled samples, our approach also adapts the general sentiment information in sentiment lexicons to target domain and incorporates it into the learning of target domain sentiment classifier. Our approach also performs better than state-of-the-art domain adaptation methods such as *SCL* (Blitzer et al., 2007) and *SFA* (Pan et al., 2010). It implies that a small number of actively selected labeled samples from target domain are beneficial for sentiment domain adaptation. *ILP* (Choi and Cardie, 2009) tries to adapt a sentiment lexicon to target domain, which is similar with our approach. *ILP* relies on labeled samples to extract the relations among words and relations between words and sentiment expressions. However, labeled samples in target domain are usually limited and the sentiment information in many unlabeled samples is not exploited in *ILP*. Thus, our approach can outperform it. Similar with our approach, *AODA* (Rai et al., 2010) and *ALCD* (Li et al., 2013) also apply active learning to domain adaptation. The major difference is that in our approach the general sentiment information extracted from sentiment lexicons is adapted to target domain, while in *AODA* and *ALCD* the sentiment classifier trained in source domains is transferred. The superior performance of our approach implies that the general sentiment information has better generalization ability than the sentiment classifier trained in a specific source domain, and is more suitable for sentiment domain adaptation.



(a) Parameter α. (b) Parameter β.

Figure 4: The influence of the parameter settings of α and β on the performance of our approach.

We further conducted several experiments to validate the advantage of our approach in training accurate sentiment classifier for target domain with only a few labeled samples. We varied the annotation budget, i.e., the number of labeled samples, from 100 to 1,000. The learning curve of our *ASDA* approach in *Book* domain is shown in Fig. 3. We also included a purely supervised sentiment classification method, i.e., *SVM*, in Fig. 3 as a baseline for comparison. Fig. 3 shows that our *ASDA* approach can consistently outperform *SVM* when the same number of labeled samples are used. The performance advantage of our approach is more significant when labeled samples are scarce. For example, the performance of our approach with only 200 labeled samples is similar to *SVM* with more than 800 labeled samples. Thus, the experimental results validate that by adapting the general sentiment information to target domain and selecting the most informative samples to annotate and learn, our approach can effectively reduce the manual annotation effort, and can train accurate sentiment classifier for target domain with much less labeled samples.

4.4 Parameter Analysis

In this section, we conducted several experiments to explore the influence of parameter settings on the performance of our approach. α and β are the two most important parameters in our approach, which control the relative importance of domain-specific sentiment similarities and the actively selected samples in training sentiment classifier for target domain. The experimental results of parameters α and β are summarized in Fig. 4.

According to Fig. 4, when α and β are too small, the performance of our approach is not optimal. This is because the useful sentiment information in domain-specific sentiment similarities mined from unlabeled samples and the actively selected labeled samples of target domain is not fully exploited. Thus, the performance of our approach improves when these parameters increase from a small value. However, when these parameters become too large, the performance of our approach starts to decline. This is because when β is too large the sentiment classifier learned by our approach is mainly decided by the limited labeled samples, and the general sentiment information extracted from sentiment lexicons is not fully exploited. When α is too large, the information in domain-specific sentiment similarities is over-emphasized, and many different words will have nearly the same sentiment weights. Thus, the performance of our approach in these scenarios is also not optimal. A moderate value of α and β is most suitable for our approach.

5 Conclusion

In this paper we present an active sentiment domain adaptation approach to train accurate sentiment classifier for target domain with less labeled samples. In our approach, the general sentiment information in sentiment lexicons is adapted to target domain with the help of a small number of labeled samples which are selected and annotated in an active learning mode. Both classification uncertainty and density are considered when selecting informative samples to label. In addition, we extract domain-specific sentiment similarities among words from unlabeled samples of target domain based on both syntactic rules and co-occurrence patterns, and incorporate them into the domain adaptation process to propagate the general sentiment information to many domain-specific sentiment words in target domain. We also propose a unified model to incorporate different types of sentiment information to train sentiment classifier for target domain. Experimental results on benchmark datasets show that our approach can train accurate sentiment classifier and at same time reduce the manual annotation effort.

Acknowledgements

This research is supported by the Key Research Project of the Ministry of Science and Technology of China (Grant no. 2016YFB0800402) and the Key Program of National Natural Science Foundation of China (Grant nos. U1536201, U1536207, and U1405254).

References

Yoshua Bengio, Olivier Delalleau, and Nicolas Le Roux. 2006. Label propagation and quadratic criterion. *Semi-supervised learning* 10.

John Blitzer, Mark Dredze, Fernando Pereira, et al. 2007. Biographies, bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification. In *ACL*. volume 7, pages 440–447. http://aclweb.org/anthology-new/P/P07/P07-1056.

Danushka Bollegala, David Weir, and John Carroll. 2011. Using multiple sources to construct a sentiment sensitive thesaurus for cross-domain sentiment classification. In *ACL:HLT*. pages 132–141. http://aclweb.org/anthology/P11-1014.

Nicolo Cesa-Bianchi, Claudio Gentile, and Luca Zaniboni. 2006. Worst-case analysis of selective sampling for linear classification. *Journal of Machine Learning Research* 7(Jul):1205–1230.

Minmin Chen, Kilian Q Weinberger, and John Blitzer. 2011. Co-training for domain adaptation. In *NIPS*. pages 2456–2464.

Yejin Choi and Claire Cardie. 2009. Adapting a polarity lexicon using integer linear programming for domain-specific sentiment classification. In *EMNLP*. pages 590–598. http://aclweb.org/anthology/D09-1062.

Yoav Freund, H. Sebastian Seung, Eli Shamir, and Naftali Tishby. 1997. Selective sampling using the query by committee algorithm. *Machine Learning* 28(2-3):133–168. http://dx.doi.org/10.1023/A:1007330508534.

Yifan Fu, Xingquan Zhu, and Bin Li. 2013. A survey on instance selection for active learning. *Knowledge and Information Systems* 35(2):249–283. https://doi.org/10.1007/s10115-012-0507-8.

Xavier Glorot, Antoine Bordes, and Yoshua Bengio. 2011. Domain adaptation for large-scale sentiment classification: A deep learning approach. In *ICML*. pages 513–520.

Mohammad Sadegh Hajmohammadi, Roliana Ibrahim, Ali Selamat, and Hamido Fujita. 2015. Combination of active learning and self-training for cross-lingual sentiment classification with density analysis of unlabelled samples. *Information sciences* 317:67–77. http://dx.doi.org/10.1016/j.ins.2015.04.003.

William L. Hamilton, Kevin Clark, Jure Leskovec, and Dan Jurafsky. 2016. Inducing domain-specific sentiment lexicons from unlabeled corpora. In *EMNLP*. pages 595–605. http://aclweb.org/anthology/D/D16/D16-1057.

Vasileios Hatzivassiloglou and Kathleen R McKeown. 1997. Predicting the semantic orientation of adjectives. In *ACL*. pages 174–181. http://aclweb.org/anthology/P/P97/P97-1023.

Yulan He, Chenghua Lin, and Harith Alani. 2011. Automatically extracting polarity-bearing topics for cross-domain sentiment classification. In *ACL:HLT*. pages 123–131. http://aclweb.org/anthology/P11-1013.

Minqing Hu and Bing Liu. 2004. Mining and summarizing customer reviews. In *KDD*. pages 168–177. http://doi.acm.org/10.1145/1014052.1014073.

Xia Hu, Lei Tang, Jiliang Tang, and Huan Liu. 2013. Exploiting social relations for sentiment analysis in microblogging. In *WSDM*. pages 537–546. http://doi.acm.org/10.1145/2433396.2433465.

Sheng Huang, Zhendong Niu, and Chongyang Shi. 2014. Automatic construction of domain-specific sentiment lexicon based on constrained label propagation. *Knowledge-Based Systems* 56:191–200. http://dx.doi.org/10.1016/j.knosys.2013.11.009.

Lianghao Li, Xiaoming Jin, Sinno Jialin Pan, and Jian-Tao Sun. 2012. Multi-domain active learning for text classification. In *KDD*. pages 1086–1094. http://doi.acm.org/10.1145/2339530.2339701.

Shoushan Li, Yunxia Xue, Zhongqing Wang, and Guodong Zhou. 2013. Active learning for cross-domain sentiment classification. In *IJCAI*. pages 2127–2133.

Bing Liu. 2012. Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies* 5(1):1–167.

Sinno Jialin Pan, Xiaochuan Ni, Jian-Tao Sun, Qiang Yang, and Zheng Chen. 2010. Cross-domain sentiment classification via spectral feature alignment. In *WWW*. ACM, pages 751–760. http://doi.acm.org/10.1145/1772690.1772767.

Sinno Jialin Pan and Qiang Yang. 2010. A survey on transfer learning. *TKDE* 22(10):1345–1359. http://dx.doi.org/10.1109/TKDE.2009.191.

Bo Pang and Lillian Lee. 2008. Opinion mining and sentiment analysis. *Foundations and trends in information retrieval* 2(1-2):1–135. http://dx.doi.org/10.1561/1500000011.

Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. 2002. Thumbs up?: sentiment classification using machine learning techniques. In *EMNLP*. pages 79–86. https://doi.org/10.3115/1118693.1118704.

Piyush Rai, Avishek Saha, Hal Daumé III, and Suresh Venkatasubramanian. 2010. Domain adaptation meets active learning. In *Proceedings of the NAACL HLT 2010 Workshop on Active Learning for Natural Language Processing*. pages 27–32. http://aclweb.org/anthology/W10-0104.

Burr Settles. 2010. Active learning literature survey. *University of Wisconsin, Madison* 52(55-66):11.

Jian Tang, Meng Qu, and Qiaozhu Mei. 2015. Pte: Predictive text embedding through large-scale heterogeneous text networks. In *KDD*. ACM, pages 1165–1174. http://doi.acm.org/10.1145/2783258.2783307.

Simon Tong and Daphne Koller. 2002. Support vector machine active learning with applications to text classification. *The Journal of Machine Learning Research* 2:45–66. http://dx.doi.org/10.1162/153244302760185243.

Peter D Turney. 2002. Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In *ACL*. pages 417–424. http://dx.doi.org/10.3115/1073083.1073153.

Leonid Velikovich, Sasha Blair-Goldensohn, Kerry Hannan, and Ryan McDonald. 2010. The viability of web-derived polarity lexicons. In *NAACL*. pages 777–785. http://www.aclweb.org/anthology/N10-1119.

Theresa Wilson, Janyce Wiebe, and Paul Hoffmann. 2005. Recognizing contextual polarity in phrase-level sentiment analysis. In *EMNLP*. pages 347–354. http://dx.doi.org/10.3115/1220575.1220619.

Fangzhao Wu and Yongfeng Huang. 2016. Sentiment domain adaptation with multiple sources. In *ACL*. pages 301–310. http://aclweb.org/anthology/P16-1029.

Fangzhao Wu, Yangqiu Song, and Yongfeng Huang. 2015. Microblog sentiment classification with contextual knowledge regularization. In *AAAI*. pages 2332–2338.

Fangzhao Wu, Sixing Wu, Yongfeng Huang, Songfang Huang, and Yong Qin. 2016. Sentiment domain adaptation with multi-level contextual sentiment knowledge. In *CIKM*. ACM, pages 949–958. https://doi.org/10.1145/2983323.2983851.

Yi Yang, Zhigang Ma, Feiping Nie, Xiaojun Chang, and Alexander G. Hauptmann. 2015. Multi-class active learning by uncertainty sampling with diversity maximization. *International Journal of Computer Vision* 113(2):113–127. http://dx.doi.org/10.1007/s11263-014-0781-x.

Dani Yogatama and Noah A. Smith. 2014. Making the most of bag of words: Sentence regularization with alternating direction method of multipliers. In *ICML*. pages 656–664.

Jingbo Zhu, Huizhen Wang, Benjamin K Tsou, and Matthew Ma. 2010. Active learning with sampling by uncertainty and density for data annotations. *IEEE Transactions on Audio, Speech, and Language Processing* 18(6):1323–1331. http://dx.doi.org/10.1109/TASL.2009.2033421.